

Taxonomy outline

From mothur

A taxonomy database consists of unaligned sequences in fasta format and a taxonomy file. The taxonomy file is a two column text file where the first column is the name of the sequence and the second column is a string of taxonomic information separated by semicolons. This information should not include spaces and the last character must be a semi-colon. For example, the first lines of silva.slv.taxonomy are as follows:

```
U87775.1      Bacteria;Alphaproteobacteria;Rhizobiales;Azorhizobium_et_rel.;Methylobacterium_et_rel.;Bosea;
DQ904772.1    Bacteria;Firmicutes;Clostridiales;Ruminococcus_et_rel.;Anaerofilum-Faecalibacterium;Faecaliba
AY553109.1    Bacteria;Firmicutes;Bacillales_Mollicutes;Bacillus_subtilis_et_rel.;Bacillus_carboniphilus_et
AY553101.1    Bacteria;Firmicutes;Bacillales_Mollicutes;Bacillus_subtilis_et_rel.;Bacillus_carboniphilus_et
```

You can download our version of the..

- SILVA reference files: The fasta and taxonomic outlines for the SILVA, greengenes, RDP, and NCBI heirarchies and can be used with the Bayesian classifier
- RDP reference files: The fasta and taxonomic outline that the RDP uses with their implementation of the Bayesian classifier
- greengenes reference files: The fasta and taxonomic outline that greengenes uses with their classifier and can be used with the Bayesian classifier

-
- This page was last modified on 8 March 2012, at 08:16.
 - Content is available under GNU Free Documentation License 1.2 unless otherwise noted.